Wald’s Test as Applied to Hypotheses in Logit Analysis

Walter W. Hauk and Allan Donner

Presented by Giles Hooker
Wald’s Tests in Logistic Regression

Consider $Y_1, \ldots, Y_N$ binary response variables from a logistic model

$$\ln \left( \frac{P(Y_i = 1|X_i)}{1 - P(Y_i = 1|X_i)} \right) = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij}$$

and consider the test of one co-ordinate

$$H_0 : \beta_k = \beta_{k0} \text{ vs } H_a : \beta_k \neq \beta_{k0}$$

In particular, we will consider two test statistics

1. Wald’s test: $T = \left( \hat{\beta}_k - \beta_{k0} \right)^2 / H_{kk}$ for $H_{kk}$ the $k, k$th entry of $(X^T WX)^{-1}$.

2. Likelihood Ratio Test:

$$L = \sum Y_i \log \left( \frac{P_i(\hat{\beta}_k)}{P_i(\beta_{k0})} \right) + (1 - Y_i) \log \left( \frac{1 - P_i(\hat{\beta}_k)}{1 - P_i(\beta_{k0})} \right)$$
A Quick Pseudo-Simulation

Two-sample comparison

- 100 in each group.
- Group 2 has 25% successes
- Move group 1 from 1% to 99%
- Test statistics for group indicator not zero.

The Wald statistics decreases at very large effect sizes.
What’s Going On?

Standard asymptotic result is that

\[ \frac{\hat{\beta}_k - \beta_k}{\sqrt{I_{kk}}} \rightarrow N(0, 1) \]

for \( I_{kk} \) the information for \( \beta_k \). Then

\[ Z = \frac{\hat{\beta}_k - \beta_k}{\sqrt{I_{kk}}} \rightarrow N \left( \frac{\beta_k - \beta_{k0}}{\sqrt{I_{kk}}} , 1 \right) \]

or \( Z^2 \) is approximately \( \chi_1^2(\xi) \) with noncentrality parameter

\[ \xi = \frac{(\beta_k - \beta_{k0})^2}{I_{kk}} \]

Since \( H_{kk} \rightarrow I_{kk} \) this applies to \( T \) as well.
What’s Going On?

Examining $\xi$:

$$I^{kk} = (I_{kk} + I_{1k}^T I_{11}^{-1} I_{k1})^{-1} \geq I_{kk}^{-1}$$

Hence

$$(\beta_k - \beta_{k0})^2 / I^{kk} \leq (\beta_k - \beta_{k0})^2 I_{kk}$$

Now

$$I_{kk} = \sum P(Y_i = 1|X_i)P(Y_i = 0|X_i)X_{ik}^2$$

and for $\beta_k \to \infty$,

$$\xi = \sum \frac{(\beta_k - \beta_{k0})^2 X_{ik}^2 e^{X_i\beta}}{(1 + e^{X_i\beta})^2} \leq \sum K_i (\beta_k - \beta_{k0})^2 e^{-|X_{ik}|\beta_k} \to 0.$$ 

Hence power goes to zero when the effect size gets large enough!
An Embarrassing Example

Data on prevalence of *T. vaginalis* in 455 college women in collected in 1976.

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Coefficient</th>
<th>Estimated S.E.</th>
<th>Wald's test</th>
<th>Likelihood ratio test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Statistic</td>
<td><em>P</em> value</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Statistic</td>
<td><em>P</em> value</td>
</tr>
<tr>
<td>1) White race</td>
<td>−4.71</td>
<td>1.66</td>
<td>8.05</td>
<td>.005</td>
</tr>
<tr>
<td>2) Vaginal discharge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>during last week</td>
<td>3.09</td>
<td>1.31</td>
<td>5.56</td>
<td>.018</td>
</tr>
<tr>
<td>3) Sexually inexperienced or partners usually use condoms</td>
<td>−10.42</td>
<td>29.81</td>
<td>0.12</td>
<td>.729</td>
</tr>
<tr>
<td>4) Smokes cigarettes</td>
<td>2.62</td>
<td>1.34</td>
<td>3.82</td>
<td>.051</td>
</tr>
<tr>
<td>5) Engages in cunnilingus</td>
<td>4.45</td>
<td>1.86</td>
<td>5.72</td>
<td>.017</td>
</tr>
<tr>
<td>6) Has ever douched</td>
<td>−3.99</td>
<td>1.73</td>
<td>5.32</td>
<td>.021</td>
</tr>
<tr>
<td>7) Colonized with <em>Mycoplasma hominis</em></td>
<td>3.98</td>
<td>1.60</td>
<td>6.19</td>
<td>.013</td>
</tr>
<tr>
<td>8) Vaginal discharge noted during physical examination</td>
<td>6.84</td>
<td>2.81</td>
<td>5.93</td>
<td>.015</td>
</tr>
<tr>
<td>9) History of gonorrhea</td>
<td>−14.46</td>
<td>107.49</td>
<td>0.02</td>
<td>.888</td>
</tr>
</tbody>
</table>

NOTE: All independent variables are dichotomous (1 = yes, 0 = no).

Particularly note discrepancies between tests for variables 3 and 9.
Conclusions

1. Suggests that Wald tests should not always be trusted.
2. Due to tail behavior of variance – true of all binary models?
3. Other exponential families? Eg, very negative parameters in Poisson case?
4. Becomes a problem for GEE models where likelihood ratio tests not available.
5. Score tests? Not addressed, not clear.
6. But, evidence occurs at extremal quantities – possibly distinguishing “true” versus “paradoxical” failures to reject = it’s bloody obvious?